ICIC Express Letters
Part B: Applications
Volume **7**, Number **11**, November **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 2353–2360

REGIONAL ANALYSIS OF FATALITY RISKS INDUCED BY NATURAL DISASTERS IN KOREA

Jangwook Oh, Teajin Lim and Moonsoo Cho

Department of Industrial and Information Systems Engineering
Soongsil University
No. 369, Sangdo-ro, Dongjak-gu, Seoul 06978, Korea
ojanguk@gmail.com; tjlim@ssu.ac.kr

Received April 2016; accepted July 2016

Abstract. *This paper analyzes regional fatality risks induced by natural disasters that have affected domestic areas of Korea from 1979 to 2013. Due to climate change, it is becoming more important to assess the risks induced by natural disasters. Extreme value distributions are considered to fit the fatality distribution, because extremely high outliers exist in the fatality data. To analyze the risks induced by natural disasters, various types of parametric models were implemented: the Weibull model, the generalized extreme value (GEV) distribution, and the generalized Pareto distribution (GPD). Next we selected the GPD for developing exceedance frequency (FN) curves. We divided the whole area of Korea into six regions, and analyzed the characteristics of natural disasters inducing fatality risk. We also estimated the 95% confidence intervals of 10-year and 30-year return levels for each region, by constructing the profile log-likelihood of the fitted GPD. The major causes of natural disasters affecting Korea include heavy rain, typhoons, storms, and heavy snow. The severity rankings of the types of natural disasters that cause the most damage are different from region to region. The results of this research will be helpful for managing regional fatality risks induced by natural disasters in Korea.*
Keywords: Exceedance frequency (FN) curve, Generalized Pareto distribution (GPD), Natural disaster, Risk analysis, Return level

1. **Introduction.** Recently, climate change has been affecting the world with more severity [1]. During the past 20 years, the earth was hit by 6,457 natural disasters including floods, typhoons and tsunamis, and over 606,000 people were killed by natural disasters and 4 billion people were affected. Over this same period, there has been an average of 335 natural disasters a year. Disasters increased 14% between 1995 and 2004 and doubled in frequency from 1985 to 1994 [2]. In Korea, an average of 20 disasters a year occurred and 800 billion Won in damages was incurred in the period from 1979 to 2013. In short, the frequency of disasters and the cost of damages are on the rise [3]. For these reasons, more attention is being paid to natural disasters in risk studies.

Historically, Korea has suffered from natural disasters such as floods and typhoons. However, a comprehensive research analyzing risks due to natural disasters has not been yet undertaken in Korea. Furthermore, the regional characteristics of natural disasters in Korea have also not yet been properly and comprehensively analyzed.

One of the oldest studies on disaster risk is Chapter 6 of the Reactor Safety Study of 1979 [4], which introduced exceedance frequency (FN) curves. After that, similar types of curves have been used in many risk studies. Recently, Lim [5] proposed the FN curves to investigate the societal risk due to natural disaster in South Korea from annually aggregated fatality data from 1916 to 2013 using EVT.

In this study, we analyze the fatality risk of natural disasters in Korea, by dividing the country into six regions. In order to investigate the total societal risk due to natural disasters, we aggregated fatality data between 1979 and 2013 from the Disaster Yearbook [3] provided by National Disaster Information Center.

We selected the FN curve as a measure for societal risk, because it provides good visual interpretation. We adopted parametric methods for fitting the FN-curve. The parametric models include the Weibull distribution, the generalized extreme value (GEV) distribution, and the generalized Pareto distribution (GPD).

Then, we selected the GPD to develop the exceedance frequency (FN) curve for each of the six regions. We analyzed the regional characteristics of natural disasters inducing fatality risk. We also estimated the 95% confidence intervals of 10-year return levels for each region, by plotting the profile log-likelihood of the fitted GPD.

This paper is organized as follows. The background to the research is described in Section 2. Statistical analysis of the regional fatality data is explained in Section 3. Development of the FN-curves by fitting the GPD and the estimation of the 10-year return levels are described in Section 4. Finally, concluding remarks are presented in Section 5.

2. Risk Analysis.

Risk denotes the combination of the probability of a hazardous event and its negative consequences, and risk analysis deals with the development of a quantitative estimate of risk. Risk analysis is an essential part of risk management, because it provides risk measures for risk assessment and risk reduction.

2.1. FN-curves.

Societal risk can be represented graphically in the FN-curve which displays the probability of exceedance as a function of the number of fatalities (N), on a double logarithmic scale as in Equation (1).

$$1 - F_N(x) = P(N > x) = \int_x^\infty f_N(y)dy \tag{1}$$

Another version of the FN-curve represents the annual exceedance frequency rather than the exceedance probability. The annual exceedance frequency can be easily obtained by multiplying the expected number of events per year by the exceedance probability in Equation (1).

2.2. Generalized Pareto distribution (GPD).

Modeling extreme or rare events is important in many areas where such events can have very negative consequences. Usually, natural disaster data have extreme values with low frequency, so extreme distributions should be considered to fit such data. Typical extreme distributions are the GEV and the GPD.

The peaks-over-threshold method deals with observations that exceed a selected threshold, and provides the GPD as the limiting distribution [7]. The GPD with a threshold u is defined as:

$$F(y) = 1 - [1 + \xi y/\{\sigma + \xi(u - \mu)\}]^{-1/\xi}, \quad y > 0 \tag{2}$$

The GPD can be represented by three extreme distributions: the Pareto ($\xi > 0$) with heavy tail, the exponential ($\xi \to 0$) with light tail, and the Beta distributions ($\xi < 0$).

2.3. Threshold selection.

Because there is bias-variance tradeoff according to the selection of the threshold, choice of an appropriate threshold is important in GPD. The following two methods with visual aids are commonly used in threshold selection.

- Mean residual life plot ⇒ The plot should be linear above the threshold.
- Parameter stability plot ⇒ The parameter estimates should be stable (i.e., constant).

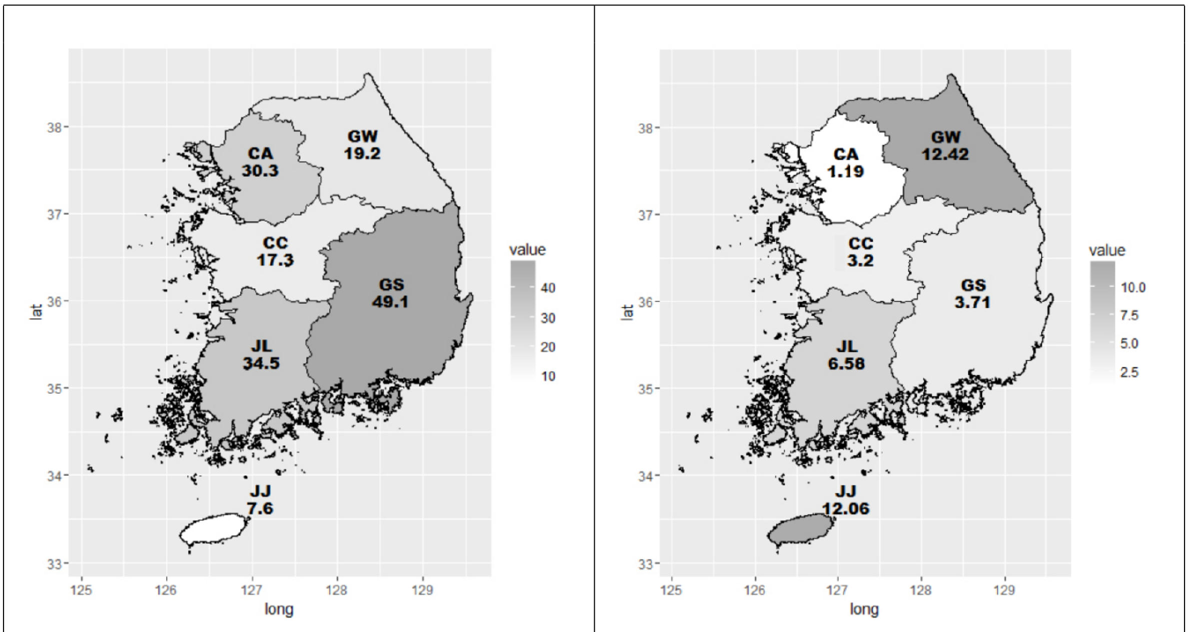
2.4. **Return level.** In extreme event analysis, it is important both when the event occurs (return period) and how big its consequences are (return level). Let ζ_u denote the probability that an observation exceeds the threshold (u), and n_y denote the number of observations per year. Then the N-year return level x_N from the GPD can be estimated by:

$$x_N = u + \{\sigma + \xi(u - \mu)\}/\xi \times \left[(N n_y \zeta_u)^{\xi} - 1\right] \tag{3}$$

3. **Data Collection and Analysis.** In this paper, we aggregated fatality data induced by natural disasters between 1979 and 2013 from the Disaster Yearbook [3] provided by National Disaster Information Center. The Disaster Yearbook provides information about the damage caused by natural disasters according to region, cause, water system and period. The major causes of natural disasters turned out to include heavy rain, typhoons, storms, heavy snow, and lightning. In order to analyze regional fatality risks, we employed R-package and library 'evd', 'maps', and 'ggplot2' [7-9].

We divided Korea into 6 areas: the capital (Seoul) area (CA), Chungcheong (CC), Gangwon (GW), Gyeongsang (GS), Jeolla (JL) and Jeju (JJ). Then we classified the frequency and fatalities according to the cause of the disaster and the region it occurred in, as in Table 1. During the time period in question, 749 natural events occurred and 5,530 people were killed. An average of close to 160 people died every year in natural disasters. The major causes were heavy rain for regions CA (84.2%) and CC (78.4%), typhoons for regions GS (59.5%), storms for JL (35.5%) and JJ (44.9%). Thus, major disasters that cause the most damage and loss of life are different from region to region.

Average total annual (individual) fatalities due to natural disasters for each region are shown in Figure 1. The highest average total annual deaths occurs in region GS (49/yr), followed by JL (34/yr), and CA (30/yr). That is, the societal risks induced by natural disasters in region GS, JL, and CA are higher than the risks in other regions. However, the highest average annual individual death risk (factor of total population size) occurs in region GW (12.42/million men-yr), followed by JJ (12.06/million men-yr), because the population sizes of region GW and JJ are smaller than those of other regions.



Average annual fatalities (/yr)	Average annual individual fatality rate (/1,000,000yr)

Figure 1. Average annual (individual) fatalities due to natural disasters for each region

TABLE 1. Frequency and fatality classified by disaster causes and regions (1979-2013)

	Frequency						**Fatality**					
	rain	**typhoon**	**storm**	**snow**	**etc.**	**sum**	**rain**	**typhoon**	**storm**	**snow**	**etc.**	**sum**
CA	80 (63.0%)	22 (17.3%)	18 (14.2%)	4 (3.1%)	3 (2.4%)	127	**892 (84.2%)**	100 (9.4%)	56 (5.3%)	8 (0.8%)	4 (0.4%)	1060
CC	59 (62.8%)	12 (12.8%)	14 (14.9%)	2 (2.1%)	7 (7.4%)	94	**475 (78.4%)**	64 (10.6%)	40 (6.6%)	2 (0.3%)	25 (4.1%)	606
GW	32 (64.4%)	13 (18.8%)	14 (20.3%)	8 (11.6%)	2 (2.9%)	69	**336 (49.9%)**	226 (33.6%)	97 (14.4%)	10 (1.5%)	4 (0.6%)	673
GS	88 (42.1%)	75 (35.9%)	37 (17.7%)	4 (1.9%)	5 (2.4%)	209	424 (24.7%)	**1021 (59.5%)**	241 (14.0%)	25 (1.5%)	6 (0.3%)	1717
JL	90 (43.9%)	38 (18.5%)	63 (30.7%)	7 (3.4%)	7 (3.4%)	205	397 (32.9%)	341 (28.3%)	**429 (35.5%)**	24 (2.0%)	16 (1.3%)	1207
JJ	6 (13.3%)	16 (35.6%)	19 (42.4%)	3 (6.7%)	1 (2.2%)	45	13 (4.9%)	111 (41.6%)	**120 (44.9%)**	21 (7.9%)	2 (0.7%)	267
sum	355 (47.4%)	176 (23.5%)	165 (22.0%)	28 (3.7%)	25 (3.3%)	749	**2537 (45.9%)**	**1863 (33.7%)**	983 (17.8%)	90 (1.8%)	57 (1.0%)	5530

4. **Risk Analysis of the Fatality Data.** For each region, we developed the exceedance frequency curve by fitting the GPD as in Figure 3.

4.1. **Choice of threshold.** To determine the threshold of the GPD, we drew a mean residual life plot and parameter stability plot.

Figure 2 illustrates those plots for region CA in the left hand side, and the diagnostic plot in the right side, respectively. Threshold for CA can be chosen at 7, because the first plot is linear above the threshold and the parameter estimates are stable at 7. Diagnostic plot also shows GPD is suited for analyzing natural disasters for CA.

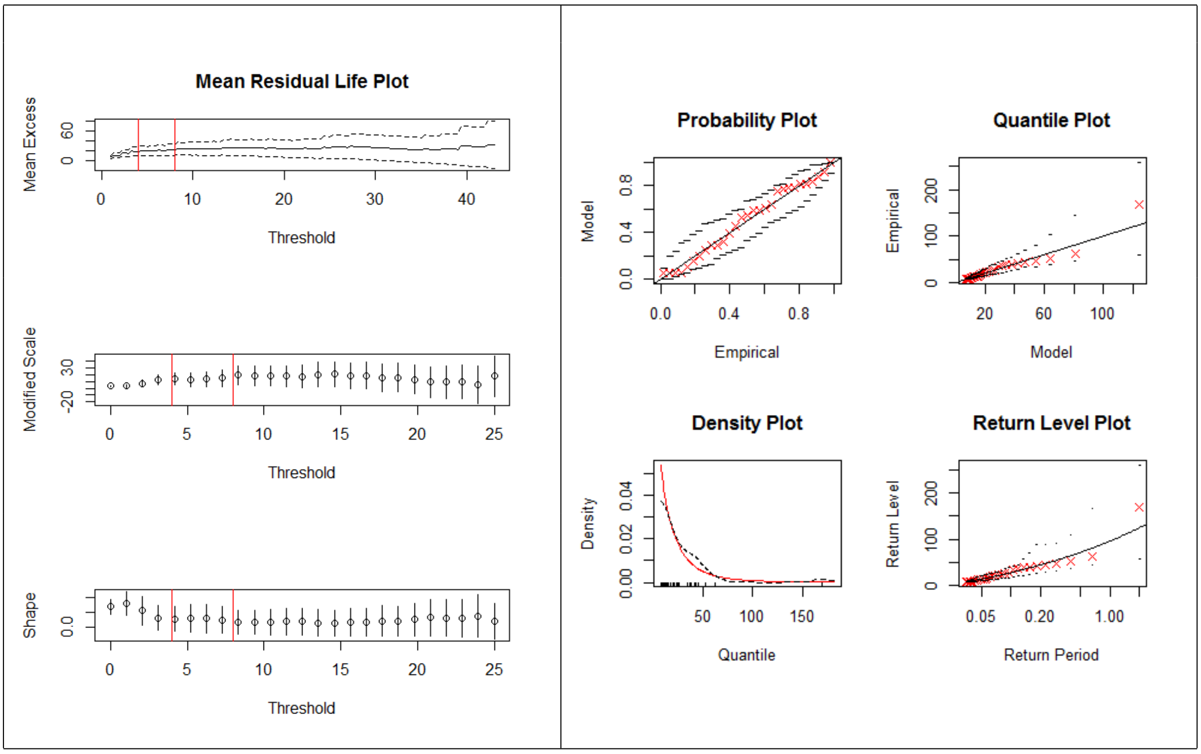


Figure 2. Choice threshold and diagnostic plot of CA

Table 2. GPD parameters in each region of Korea

	CA	CC	GS	GW	JL	JJ
Threshold (μ)	7	4	7	5	4	4
Scale (σ)	17.1551	4.9851	11.0591	11.6158	6.0843	7.0285
Shape (ε)	0.2386	0.7499	0.4538	0.3302	0.2493	0.1177

4.2. **Parameter estimation.** The estimated parameters of the GPD by maximum likelihood method are shown in Table 2. The positive value of shape parameters in each region implies heavy-tailed distribution. The FN-curves from the GPD models are shown in Figure 3. They look fine in the middle area, but some of them show lack of fit in the tail area. This is due to an insufficient amount of data.

4.3. **Return levels for 10- and 30-year period.** Due to the lack of fit in the tail area, it is not very much meaningful to estimate long year return levels, so we estimated 10-year return levels as shown in Table 3 and Figure 4. The right hand side of Figure 4 illustrates the profile log-likelihood for region CA to estimate the 95% confidence interval.

We also estimated 30-year return levels as shown in Table 4 and Figure 5. The right hand side of Figure 5 illustrates the profile log-likelihood for region CA to estimate the 95% confidence interval.

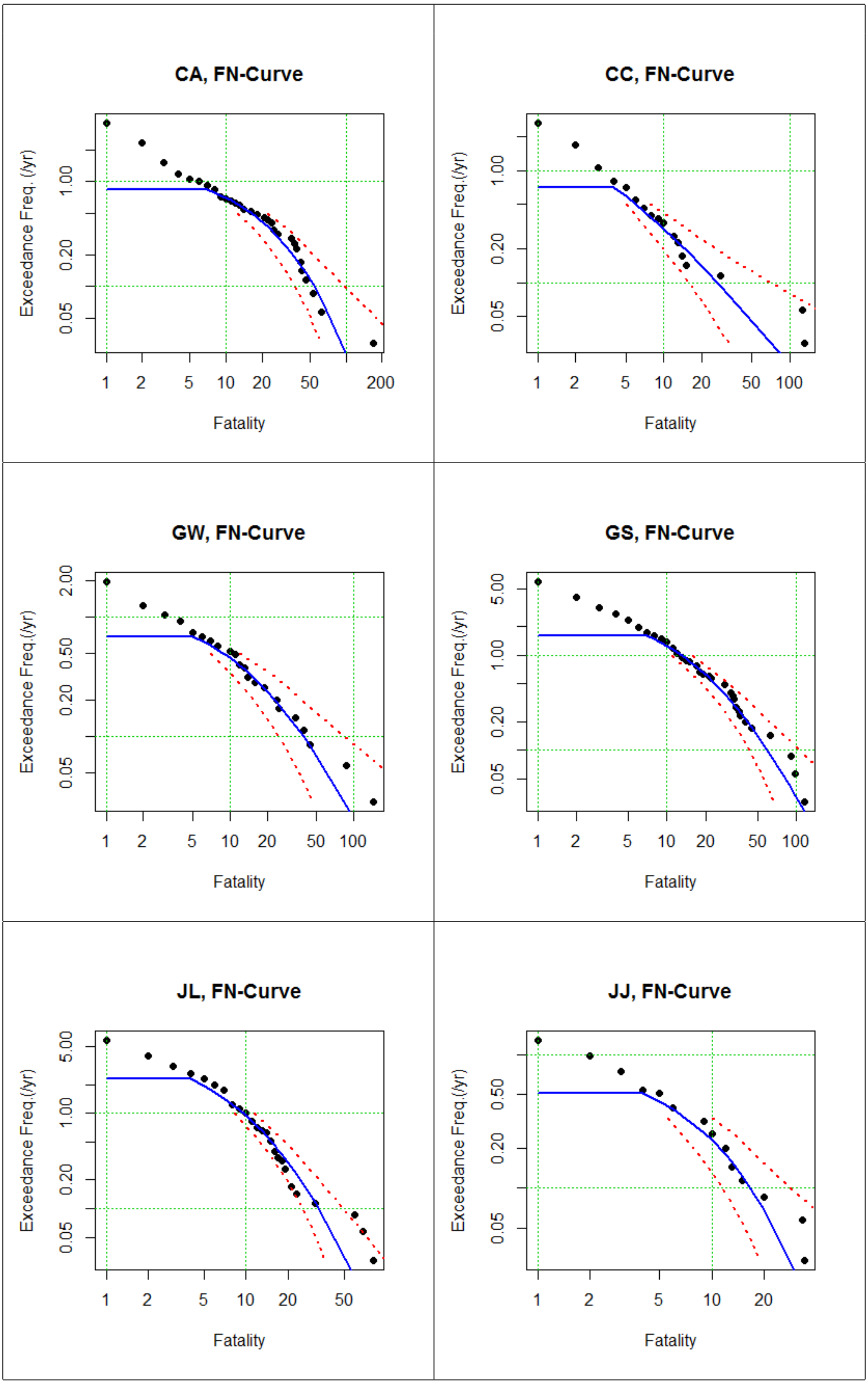


Figure 3. FN-curve of GPD in each region of Korea

Table 3. Return level for 10-year period in 6 regions in Korea (CL: 95%)

	CA	CC	GS	GW	JL	JJ	Total
Estimate	55.38	27.02	60.29	39.87	33.23	16.95	232.74
Lower	38.31	14.04	43.98	25.3	25.56	11.7	159.09
Upper	100.28	72.94	108.21	86.74	50.24	29.01	447.42

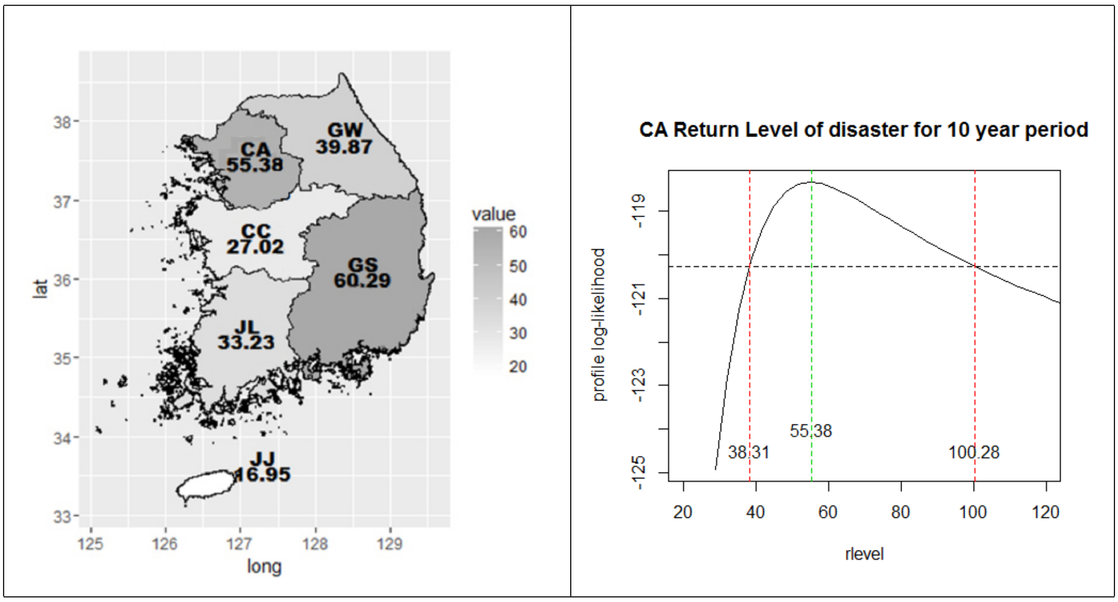


Figure 4. Map of 10-year return level and the profile log-likelihood of CA

Table 4. Return level for 30-year period in 6 regions in Korea (CL: 95%)

	CA	CC	GS	GW	JL	JJ	Total
Estimate	90.15	64.18	97.92	77.12	49.98	26.96	406.31
Lower	59.44	30.52	64.54	44.03	35.67	18.09	252.29
Upper	266.13	401.18	254.80	323.21	93.38	80.22	1418.92

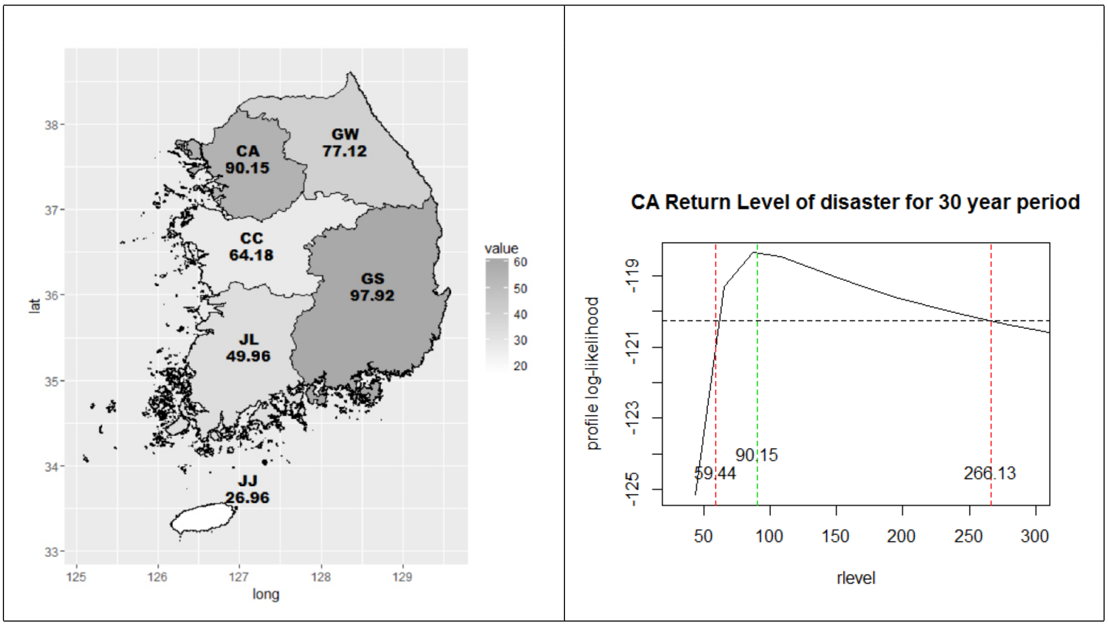


Figure 5. Map of 30-year return level and the profile log-likelihood of CA

5. **Conclusions.** In this study, we analyzed the fatality data concerning natural disasters in six regions of Korea from 1979 to 2013. According to the frequency and consequence of disasters, the most common cause of severe damage was heavy rain, followed by typhoons and storms. Heavy snow and other natural disasters revealed relatively less impact.

The major causes were heavy rain for regions CA and CC, typhoons for region GS and JJ, and storms for JL, respectively. In light of this variation, each region needs to prepare for the specific type of disaster that affects the region most. As for the individual fatality rates, regions GW and JJ are the most vulnerable to disasters, and these regions require even stronger risk management plans for natural disasters.

The 30-year return levels estimated in the 6 areas show that we expect 406 fatalities due to natural disasters for the next 30-year period, and at least 252 fatalities with 97.5% confidence. In light of this situation, Korea needs to pay much attention to the risk management of natural disasters.

Further studies will be required to use mixture distribution, because it would be more suitable for fitting natural disasters. Also, future work is needed to collect more data prior to 1988 to obtain better results.

Acknowledgment. This work was supported by National Research Foundation of Korea. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the article.

REFERENCES

[1] IPCC, Climate change 2014: Impacts, adaptation, and vulnerability, *Working Group II Contribution to the 5th Assessment Report of the Intergovernmental Panel on Climate Change*, http://www.ipcc.ch, 2014.
[2] UNISDR, *The Human Cost of Weather Related Disasters 1995-2015*, https://www.unisdr.org, 2015.
[3] NEMA (National Emergency Management Agency), *Yearbook of Natural Disaster*, 1979-2014.
[4] WASH-1400, *Reactor Safety Study*, Chapter 6, 1979.
[5] T.-J. Lim, Risk analysis of natural disaster in South Korea using extreme value theory, *ICIC Express Letters, Part B: Applications*, vol.7, no.2, pp.301-306, 2016.
[6] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, 2001.
[7] A. Stephenson, *Package 'evd'*, Manual, pp.1-87, 2015.
[8] B. Ray, *Package 'maps'*, Manual, pp.1-36, 2016.
[9] W. Hadley, C. Winston and R. Studio, *Package 'ggplot2'*, Manual, pp.1-197, 2016.